FEBS Letters 586 (2012) 2841–2845

FEBS Letters

journal homepage: www.FEBSLetters.org

Review

Molecular basis for population variation: From SNPs to SAPs

Jia-Rui Wu *, Rong Zeng *

Key Laboratory of Systems Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

ARTICLE INFO

Article history:
Received 16 June 2012
Revised 14 July 2012
Accepted 16 July 2012
Available online 22 July 2012

Edited by Miguel De la Rosa, Felix Wieland and Wilhelm Just

Keywords:
SNP
Single amino-acid polymorphism (SAP)
RNA editome
Protein editing
The Central Dogma

ABSTRACT

Single nucleotide polymorphisms (SNPs) are one type of genomic DNA variations in a population. Correspondingly, single amino-acid polymorphisms (SAPs) derived from non-synonymous SNPs represent protein variations in a population. Recently, using proteomic approaches, SAPs in the plasma proteomes of an Asian population were systematically identified for the first time. That study showed that heterozygous and homozygous proteins with various SAPs have different associations with particular traits in the population. Recent discoveries of widespread differences between RNA and DNA sequences indicate that RNA editing is also a source of SAPs – one that is independent of genomic SNPs. Furthermore, we argue that there are de novo SAPs that are not encoded by either DNA or RNA sequences.

1. Introduction

Since the 2003 completion of the Human Genome Project, which provided sequencing data at the species level, the scientific community has concentrated on genetic variants at the population level to understand and develop treatments for common complex diseases such as cancer and diabetes. In the same year, the International HapMap Project was initiated; this project seeks to identify common patterns of DNA sequence variation in the genomes of human populations with ancestry in parts of Africa, Asia and Europe [1]. Single nucleotide polymorphisms (SNPs), resulting from single base variations, are recognised as the most common type of genetic variants in the human genome; these variants are often associated with particular physiological or pathological traits in individuals. It has been estimated that the human genome contains approximately 10 million SNPs [2]. In 2005, the International HapMap Consortium released the first version of the haplotype map of the human genome, in which more than one million SNPs were analysed [3]; the second version, released in 2007, involved more than three million SNPs [4].

Currently, millions of SNPs are available for physiological or pathological association studies. In general, there are two types of SNPs: functional SNPs, such as those altering an amino sequence or a transcription factor binding element, and non-functional SNPs, which are not related to the regulation of protein expression. To avoid the potential problem associated with the use of non-functional SNPs in association studies, which might result in different degrees of linkage disequilibrium with functional SNPs among different populations [5], Humphries et al. strongly recommended using functional SNPs in association studies [6]. Therefore, it is necessary to analyse functional SNPs at the proteome level.

2. Molecular variation in populations from non-synonymous SNPs to single amino-acid polymorphisms

SNPs can be located either within coding regions or outside coding regions. Because of the degeneracy of the genetic code, SNPs within coding regions may not change the amino acid sequence of the corresponding proteins. Therefore, SNPs within coding regions are further divided into two categories: synonymous SNPs and non-synonymous SNPs. Non-synonymous SNPs belong to the previously defined functional category. Recently, Schaefer et al. designed an extensive database of non-synonymous SNPs, called the SNP database of effects (SNPdbe), which contains 1.3 million total SNPs corresponding to single amino acid substitutions from three SNP databases [7].

Several reports have shown that non-synonymous polymorphisms resulting in changes in amino acid sequences often have significant impacts on gene functions [8–10]. Yoshiura et al. showed that SNP rs17822931, a non-synonymous variant (180G/R) in the *ABCC11* gene, was the sole determinant of the human earwax type, whereas the SNPs located in the intron and Alu-repetitive sequence did not have similar effects on this physiological

* Corresponding authors.
E-mail addresses: wujr@sibs.ac.cn (J.-R. Wu), zr@sibs.ac.cn (R. Zeng).

0014-5793 © 2012 Federation of European Biochemical Societies. Published by Elsevier B.V. Open access under CC BY-NC-ND license.
http://dx.doi.org/10.1016/j.febslet.2012.07.036

trait, even though these SNPs had the same *P* values as the SNP rs17822931 in the association study [8]. In a case-control study with 52 608 gene-based tag SNPs selected from the JSNP database, Kubo et al. determined that the non-synonymous SNP rs2230500 (1425G/A) in exon 9 of the gene *PRKCH*, a member of protein kinase C family, can increase the risk of cerebral infarction [9]. A recent case-control study with a Chinese cohort and biochemical assay showed that the non-synonymous SNP rs231775 (49G/A) in the gene *CTLA-4*, which is involved in the regulation of T-cell activation, was associated with increased susceptibility to multiple cancers, including lung, breast, oesophageal and gastric cancers [10]. In contrast, some non-synonymous SNPs do not affect the functions of genes. For example, the non-synonymous SNP rs34952165 (122 V/I) in the DNase 1L1 gene was shown to have little effect on the DNase activity of the enzyme [11].

To understand the genotype distributions of non-synonymous SNPs at the population level, Takeshita et al. systematically analysed 13 non-synonymous SNPs in three apoptosis-related endonuclease genes, ENDOG, DFFB and FEN-1, in 13 different human populations of healthy Asians, Africans and Caucasians [12]. These researchers showed that only three non-synonymous SNPs, including S12L in the ENDOG gene and R196K and K277R in the DFFB gene, are polymorphic, whereas the other 10 SNPs are not [12]. Based on their previous works [13–15], Takeshita et al. concluded that endonucleases exhibit a low degree of genetic heterogeneity with regard to non-synonymous SNPs [12]. By re-sequencing the human exomes of 200 individuals from Denmark, Li et al. identified 53 081 SNPs within the coding regions of 18 654 well-annotated genes; half of the identified coding SNPs were non-synonymous [16]. These researchers further revealed that there were twice as many deleterious non-synonymous SNPs as synonymous SNPs in the low-frequency range, with minor allele frequencies of 2–5%, implying that low-frequency non-synonymous SNPs can have a substantial impact on evolutionary fitness [16].

Because non-synonymous SNPs within coding regions must cause amino-acid variation in proteins during the translational process, we propose that the specific properties of these amino acid variations can only be uncovered by direct analysis at the protein level. Although SNPs have been studied extensively over the past decade, little attention has been paid to the amino-acid polymorphisms of proteins at the population level. A search of the PubMed database using the term "SAAP" (single amino acid polymorphism) found only four papers focused on SAAPs [17–20], of which two papers reported SAAPs identified by the analysis of individual protein and vascular tissues, respectively, using proteomic approaches [17,18], and the other two papers provided SAAP databases generated with bioinformatics tools [19,20]. Recently, we performed the first analysis of single amino-acid polymorphisms in plasma proteins at the population level, and coined the term "SAP" to be simpler than "SAAP" and to fit the corresponding term "SNP" [21].

To systematically detect SAPs in human plasma based on high-confidence Swiss-Prot annotation, we first constructed a reliable plasma-related SAP database containing 439 SAPs matched to 29 plasma proteins, and then we analysed the plasma proteomes of 33 plasma samples from Asian individuals using shotgun proteomics [21–23]. We identified 2154 SAP peptides corresponding to 46 unique SAP peptides [21]. To further uncover the properties of SAPs at the population level, three pairs of SAP peptides corresponding to complement components C5 and C7 and complement factor H (CFH) were selected among the identified SAP peptides for quantitative analysis in plasma samples from 290 Asian individuals using a targeted proteomic approach (selected reaction monitoring, SRM), in which the target protein samples were mixed with stable-isotope labelled peptides and then selectively quantified by mass spectrometry [24–26].

3. Qualitative and quantitative analysis of SAPs in plasma from Asian individuals at the population level

For humans and other diploid-genome organisms, an individual can be either homozygous or heterozygous for a particular SNP [27,28]. It has been reported that the different genetic structures of coding SNPs are tightly associated with the physiological and pathological traits of individuals [8–10,29]. For example, Yoshiura et al. reported that AA-homozygosity of the non-synonymous SNP rs17822931 in the *ABCC11* gene in Japanese individuals is associated with dry earwax, whereas GG-homozygosity and GA-heterozygosity are associated with wet earwax [8]. In addition, the frequency of the homozygous *CTLA-4* 49AA (non-synonymous SNP rs231775) genotype was significantly higher in Chinese patients with lung cancer, whereas the heterozygous *CTLA-4* 49GA genotype was not associated with lung cancer [10]. Dr. Chess' group at Harvard Medical School reported widespread monoallelic expression from human autosomes [30], suggesting that some SNPs could be distributed with no genetic heterogeneity in the population. In the case of the *FEN-1* gene, all six non-synonymous SNPs have been detected as entirely monoallelic in the 13 examined populations [12].

By analysing these three pairs of SAP peptides with the SRM approach, we showed that the percentage of diabetic and overweight/obese Chinese individuals heterozygous for the C5 SAPs was significantly higher than normal individuals, whereas the distributions of the heterozygotes for C7 and CFH SAPs were not significantly different in these Asian sub-populations [21, Fig. 2A]. We further showed that the proportion of C7_587P homozygosity in normal individuals was significantly higher than that in individuals with a metabolic disorder [21, Fig. 2B]. Taken together, these results suggest that the genetic structures of SAPs at the proteome level have distribution properties similar to those of SNPs at the genome level in populations.

In heterozygous individuals, allele-specific expression of SNPs is widely detected at the transcriptome level [31–33]. Approximately 11–22% of heterozygous mRNA-associated SNPs exhibited allele-specific expression in each of four human cell lines, and 4.3–8.5% of allele-specific expression was tissue specific [34]. This type of allele-specific expression of heterozygous mRNA-SNPs is often associated with physiological and pathological traits in individuals. A recent report showed that H2-haplotype individuals with *TGFBR1* allele-specific expression were highly susceptible to colorectal cancer [35]. Although it is possible to obtain quantitative information regarding the allele-specific expression of heterozygous SNPs at the transcriptional level, understanding polymorphism-related expression at the proteome level requires a quantitative proteomic approach. Recently, with a newly developed fluorescent-protein detection system, Dr. Xie's group at Harvard University quantitatively analysed the correlation between protein and mRNA expression at single-molecule level and concluded that the protein and mRNA copy numbers for any given gene are uncorrelated [36]. This work further strengthens our view that the quantitative analysis of SAPs at the proteome level is necessary to fully understand the quantitative properties of genetic polymorphisms, which cannot be interpreted simply based on the allele-specific expression of mRNA-SNPs at the transcriptome level.

Taking advantage of the SRM approach [24–26], we analysed each SAP peptide of these three pairs of SAPs in the plasma of these 290 Asians in terms of the absolute concentration [21]. This quantitative analysis showed that three SAP peptides (C7_587T, CFH_62 V and CFH_62I) in the homozygous state were significantly associated with metabolic disorders [21, Fig. 3]. We quantitatively measured the distributions of the three pairs of SAPs in the heterozygous state in all four sub-populations of the examined Asians

and revealed various associations of these heterozygous SAPs with particular physiological and pathological traits [21, Fig. 4].

Our recent study is the first to reveal the quantitative properties of protein-polymorphisms at the population level. We draw three conclusions from this study: (1) Traditional methods for the analysis of protein expression, such as Western-blotting and ELISA, which can detect the total concentration of a protein, are unable to separately measure the absolute concentration of each SAP protein in heterozygous individuals, but SRM provides a useful tool for performing this analysis. (2) Each counterpart-SAP of a heterozygous SAP-pair might present differential distributions associated with physiological or pathological traits, e.g., C7_587T in the heterozygotes is associated with obesity and diabetes, whereas C7_587P is not [21, Fig. 4]. Our recent work analysing more than one thousand Asian plasma samples using SRM approach revealed that one SAP peptide of the heterozygous SAP peptide pair in Apoa4 is associated with lower triglyceride levels, whereas its counterpart is associated with higher triglyceride levels (unpublished data). (3) The SRM-based analysis of plasma SAPs, particularly heterozygous polymorphisms, might uncover new biomarkers associated with particular physiological or pathological traits.

4. Three possible SAP sources: a challenge to the Central Dogma

Originally, the concept of SAPs was derived from the concept of SNPs on the basis of the Central Dogma of molecular biology (Fig. 1), which was first proposed by Crick F. in 1958 [37]. The Central Dogma describes how genetic information is transferred from the genome level to the protein level, i.e., the information in gene sequences is delivered to the nucleic acids of mRNA molecules through transcription and is then transferred to the amino acids of proteins according to genetic code through translation. Based on the Central Dogma, it is a long-standing belief that mRNA sequences must be faithful to DNA sequences and that the amino acid sequence of the resultant proteins must be faithfully determined by the genetic code. Therefore, genomic non-synonymous SNPs in coding regions must be the major source of SAPs according to the Central Dogma.

Although the "mRNA editing" phenomenon that results in nucleotide variants of mRNA molecules during the transcriptional process was discovered in the early 1980s [38], this phenomenon has not destroyed the fundamental belief in the Central Dogma and is recognised as a rare variation in the genetic information at the level of transcription. However, Cheung's group at the University of Pennsylvania recently showed there are widespread RNA and DNA sequence differences in the human transcriptome; more than 10000 exonic sites did not match the DNA sequences, and all 12 possible categories of nucleic-acid discordances were observed [39]. Although some scientists argued that Cheung and his colleagues may have generated a large number of false positives during their analysis of the sequencing data [40,41], and Schrider et al. even came to the opposite conclusion by re-analysing the Cheung's dataset [42], more and more studies have provided strong evidence to support Cheung's conclusion. Using siRNA-knockdown of the ADAR protein, which is a well-known RNA-editing enzyme that catalyses the transformation of adenosine to inosine [43], Bahn et al. showed that genome-wide DNA–RNA differences were significantly inhibited, indicating that the extensive nucleotide variants at the transcription level are generated through RNA editing mechanisms. This collection of edited RNAs was then called the "RNA editome" [44]. Bahn et al. further indicated that the majority of RNA editing sites belonged to the A-to-I type, and genes with predicted A-to-I editing were significantly enriched in the category of cancer-related genes, suggesting the possible existence of cancer-specific RNA editing [44]. Another recent report further supported Bahn's conclusions and even extended the existence of the RNA editome to microRNA molecules [45]. Snyder and his colleagues showed for the first time that many RNA edits in peripheral blood mononuclear cells that were continuously collected from a healthy individual over 14 months exhibited dynamic and differential changes during that period, suggesting that the components of the RNA editome could be dependent on the environment [46]. The discovery of the editome challenges the Central Dogma: there might be extensive nucleotide variation introduced during the transmission of the genetic information from the genome to the transcriptome. As Hayden notes, "If verified, the findings would require a rewrite of the 'central dogma' of molecular biology" [41].

In this regard, we propose that RNA editome must be a second source of SAPs, which is independent on the source of SNPs (Fig. 1). By globally analysing proteins in human B cells using a proteomic approach, Cheung and his colleagues concluded that, in human cells, there are peptide sequences that are not exactly encoded by the DNA coding sequences [39]. In addition, by comparing the mass spectra of proteins to the amino acid sequences of proteins with 4586 missense single nucleotide variants (equivalent to non-synonymous SNPs) and to all 30385 RNA edits, Snyder and colleagues identified peptides corresponding to 48 single nucleotide variants and 51 edits and concluded that several transcripts containing RNA edits serve as templates for protein synthesis [46].

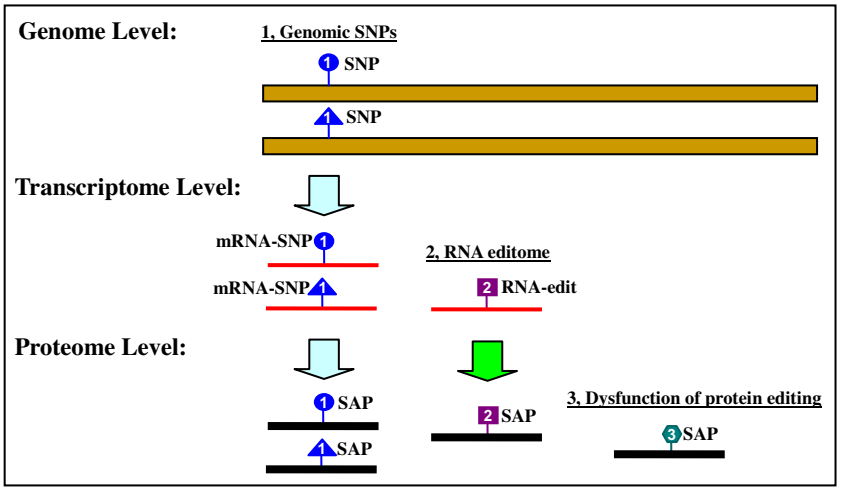

Fig. 1. Three possible sources of SAPs. Schematic illustration of three possible sources of SAPs. No. 1 indicates the major SAP-source from genomic non-synonymous SNPs; No. 2 indicates a RNA-editing dependent SAP-source; and No. 3 indicates a possible SAP-source due to dysfunction of protein editing, which is independent on DNA/RNA sequences.

Furthermore, other RNA-modification mechanisms, such as RNA oxidation and pseudouridylation, that could cause transcriptional errors are also possible sources of SAPs at the transcriptome level [47–49].

In contrast to the meaning of RNA editing that causes nucleotide variation, a term of protein editing means a proofreading of amino acids during the process of protein synthesis [50,51]. As early as the 1970s, researchers found that aminoacyl-tRNA (aa-tRNA) synthetases could destroy mislabelled amino acids and hydrolyse mischarged tRNAs in the editing domain of the aa-tRNA synthetase [52]. Subsequently, the structural details of the editing domain were analysed systematically [reviewed by 53,54]. Recently, another protein-editing mechanism has been uncovered wherein the ribosome selects correct aa-tRNAs and disregards the incorrect ones using a so-called induced-fit mechanism [55]. Theoretically, dysfunctional protein editing should result in the loss of fidelity during translation. Lee et al. reported that an editing-defective aa-tRNA synthetase resulted in an increase in the number of mischarged tRNAs, which could result in the intracellular accumulation of misfolded proteins in neurons [56]. By introducing a transgene expressing an editing-deficient aa-tRNA synthetase into mammalian cells, Nangle et al. detected an increase in the misincorporation of a particular amino acid into an EGFP-reporter [57]. A higher translational error rate due to the suppression of protein editing functions might have a selective advantage during evolution. Li et al. recently showed that the parasite *Mycoplasma* naturally has point mutations and deletions in the editing domains of its aa-tRNA synthetases, resulting in mistranslations of the parasite proteins [58]. These researchers proposed that these natural protein-editing mutations in the parasites might have evolved as a mechanism to generate antigen diversity to escape host defence systems [58].

It has been estimated that the typical frequencies of mistranslation are approximately 1 per $10^3 - 10^4$ codons [54,59]. Based on this error rate, 18% of proteins having an average length of 400 amino acids contain at least one missense substitution [60]. Because the analysis of amino-acid variants at the proteome level has been less extensive than analysis of the RNA editome at transcriptome level so far, the predicted protein variation rate is most likely underestimated, suggesting that there might be widespread amino-acid variants in a population. Therefore, we propose a third source of SAPs, which is independent of SNPs and the RNA editome, resulting from the dysfunction of protein editing or other unknown mechanisms related to protein synthesis or post-protein synthesis processes (Fig. 1). In other words, cells might contain de novo SAPs that are independent of DNA/mRNA sequences, an idea that again challenges the Central Dogma at the proteome level.

5. Future perspectives

In light of the recent discovery of the RNA editome, it is likely that RNA editing is a second source of SAPs, which is independent of non-synonymous SNPs that are a major source of SAPs. Furthermore, we propose a third source of SAPs: dysfunctional protein editing or other unknown chemical amino-acid modifications lead to de novo SAPs, which is independent of nucleotide sequences and should form a new area of population variation research (Fig. 1).

The study of SAPs should provide new insight into the Central Dogma and the relationship between genotype and phenotype. The study of SAPs could also lead to the identification of new molecular biomarkers that can be used to help treat complex diseases and improve human health.

Acknowledgements

We apologise to the scientists whose work could not be included in this mini-review because of space limitations. We are grateful for the related funding support from the Ministry of Science and Technology (2011CB910200, 2011CB910601) and the National Natural Science Foundation of China (31130034).

References

[1] The International Hap Map Consortium (2003) The International HapMap project. Nature 426, 789–796.
[2] Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Shaw, N., Lane, C.R., Lim, E.P., Kalyanaraman, N., et al. (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes. Nat. Genet. 22, 231–238.
[3] The International HapMap Consortium (2003) A haplotype map of the human genome. Nature 437, 1299–1320.
[4] The International HapMap Consortium (2007) A second generation human haplotype map of over 3.1 million SNPs. Nature 449, 851–861.
[5] Neale, B.M. and Sham, P.C. (2004) The future of association studies: gene-based analysis and replication. Am. J. Hum. Genet. 75, 353–362.
[6] Humphries, S.E., Ridker, P.M. and Talmud, P.J. (2004) Genetic testing for cardiovascular disease susceptibility: a useful clinical management tool or possible misinformation? Arterioscler. Thromb. Vasc. Biol. 24, 628–636.
[7] Schaefer, C., Meier, A., Rost, B. and Bromberg, Y. (2012) SNPdbe: constructing an nsSNP functional impacts database. Bioinformatics 28, 601–602.
[8] Yoshiura, K., Kinoshita, A., Ishida, T., Ninokata, A., Ishikawa, T., Kaname, T., Bannai, M., Tokunaga, K., Sonoda, S., Komaki, R., et al. (2006) A SNP in the ABCC11 gene is the determinant of human earwax type. Nat. Genet. 38, 324–330.
[9] Kubo, M., Hata, J., Ninomiya, T., Matsuda, K., Yonemoto, K., Nakano, T., Matsushita, T., Yamazaki, K., Ohnishi, Y., Saito, S., et al. (2007) A non-synonymous SNP in PRKCH (protein kinase Cη) increases the risk of cerebral infarction. Nat. Genet. 39, 212–217.
[10] Sun, T., Zhou, Y., Yang, M., Hu, Z., Tan, W., Han, X., Shi, Y., Yao, J., Guo, Y., Yu, D., et al. (2008) Functional genetic variations in cytotoxic T-lymphocyte antigen 4 and susceptibility to multiple types of cancer. Cancer Res. 68, 7025–7034.
[11] Ueki, M., Fujihara, J., Takeshita, H., Kimura-Kataoka, K., Iida, R., Nakajima, T., Kominato, Y., Yuasa, I. and Yasuda, T. (2010) Genetic and expression analysis of all non-synonymous single nucleotide polymorphisms in the human deoxyribonuclease I-like 1 and 2 genes. Electrophoresis 31, 2063–2069.
[12] Takeshita, H., Fujihara, J., Ueki, M., Iida, R., Koda, Y., Soejima, M., Yuasa, I., Kato, H., Nakajima, T., Kominato, Y. and Yasuda, T. (2012) Non-synonymous single-nucleotide polymorphisms of the human apoptosis-related endonuclease – DNA fragmentation factor beta polypeptide, endonuclease G, and flap endonuclease-1 – genes show a low degree of genetic heterogeneity. DNA Cell Biol. 31, 36–42.
[13] Ueki, M., Takeshita, H., Fujihara, J., Iida, R., Yuasa, I., Kato, H., Panduro, A., Nakajima, T., Kominato, Y. and Yasuda, T. (2009) Caucasian-specific allele in non-synonymous single nucleotide polymorphisms of the gene encoding deoxyribonuclease I-like 3, potentially relevant to autoimmunity, produces an inactive enzyme. Clin. Chim. Acta 407, 20–24.
[14] Yasuda, T., Ueki, M., Takeshita, H., Fujihara, J., Kimura-Kataoka, K., Iida, R., Tsubota, E., Soejima, M., Koda, Y., Kato, H. and Panduro, A. (2010) A biochemical and genetic survey on all non-synonymous single nucleotide polymorphisms of the gene encoding human deoxyribonuclease I potentially relevant to autoimmunity. Int. J. Biochem. Cell Biol. 42, 1216–1225.
[15] Fujihara, J., Ueki, M., Yasuda, T., Iida, R., Soejima, M., Koda, Y., Kimura-Kataoka, K., Kato, H., Panduro, A., Tongu, M. and Takeshita, H. (2011) Functional and genetic survey of all known single-nucleotide polymorphisms within the human deoxyribonuclease I gene in wide-ranging ethnic groups. DNA Cell Biol. 30, 205–217.
[16] Li, Y., Vinckenbosch, N., Tian, G., Huerta-Sanchez, E., Jiang, T., Jiang, H., Albrechtsen, A., Andersen, G., Cao, H., Korneliussen, T., et al. (2010) Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. Nat. Genet. 42, 969–972.
[17] Valentine, S.J., Sevugarajan, S., Kurulugama, R.T., Koeniger, S.L., Merenbloom, S.I., Bohrer, B.C. and Clemmer, D.E. (2006) Split-field drift tube/mass spectrometry and isotopic labeling techniques for determination of single amino acid polymorphisms. J. Proteome Res. 5, 1879–1887.
[18] Abraham, P., Adams, R., Giannone, R.J., Kalluri, U., Ranjan, P., Erickson, B., Shah, M., Tuskan, G.A. and Hettich, R.L. (2012) Defining the boundaries and characterizing the landscape of functional genome expression in vascular tissues of Populus using shotgun proteomics. Proteome Res. 11, 449–460.
[19] Cavallo, A. and Martin, A.C. (2005) Mapping SNPs to protein sequence and structure data. Bioinformatics 21, 1443–1450.
[20] Hurst, J.M., McMillan, L.E., Porter, C.T., Allen, J., Fakorede, A. and Martin, A.C. (2009) The SAAPdb web resource. A large-scale structural analysis of mutant proteins. Hum. Mutat. 30, 616–624.
[21] Su, Z., Sun, L., Yu, D., Li, R., Li, H., Yu, Z., Sheng, Q., Lin, X., Zeng, R. and Wu, J. (2011) Quantitative detection of single amino-acid polymorphisms by targeted proteomics. J. Mol. Cell Biol. 3, 309–315.
[22] Sheng, Q., Dai, J., Wu, Y., Tang, H. and Zeng, R. (2012) Build summary: using a group-based approach to improve the sensitivity of peptide/protein identification in shotgun proteomics. J. Proteome Res. 11, 1494–1502.
[23] Shao, C., Liu, Y., Ruan, H., Li, Y., Wang, H., Kohl, F., Goropashnaya, A.V., Fedorov, V.B., Zeng, R., Barnes, B.M. and Yan, J. (2010) Shotgun proteomics analysis of hibernating arctic ground squirrels. Mol. Cell. Proteomics 9, 313–326.

[24] Addona, T.A., Abbatiello, S.E., Schilling, B., Skates, S.J., Mani, D.R., Bunk, D.M., Spiegelman, C.H., Zimmerman, L.J., Ham, A.L., Keshishian, H., et al. (2009) Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. Nat. Biotechnol. 27, 633–641.
[25] Picotti, P., Bodenmiller, B., Mueller, L.N., Domon, B. and Aebersold, R. (2009) Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. Cell 138, 795–806.
[26] Picotti, P., Rinner, O., Stallmach, R., Dautel, F., Farrah, T., Domon, B., Holger Wenschuh, H. and Aebersold, R. (2010) High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. Nat. Methods 7, 43–46.
[27] Pastinen, T. (2010) Genome-wide allele-specific analysis: insights into regulatory variation. Nat. Rev. Genet. 12, 533–538.
[28] Tewhey, R., Bansal, V., Torkamani, A., Topol, E.J. and Schork, N.J. (2011) The importance of phase information for human genomics. Nat. Rev. Genet. 12, 215–223.
[29] Wiechec, E. and Hansen, L.L. (2009) The effect of genetic variability on drug response in conventional breast cancer treatment. Eur. J. Pharm. 625, 122–130.
[30] Gimelbrant, A., Hutchinson, J.N., Thompson, B.R. and Chess, A. (2007) Widespread monoallelic expression on human autosomes. Science 318, 1136–1140.
[31] Yan, H., Yuan, W., Velculescu, V.E., Vogelstein, B. and Kinzler, K.W. (2002) Allelic variation in human gene expression. Science 297, 1143.
[32] Pickrell, J.K., Marioni, J.C., Pai, A.A., Degner, J.F., Engelhardt, B.E., Nkadori, E., Veyrieras, J., Stephens, M., Gilad, Y. and Pritchard, J.K. (2010) Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature 464, 768–772.
[33] Montgomery, S.B., Sammeth, M., Gutierrez-Arcelus, M., Lach, R.P., Ingle, C., Nisbett, J., Guigo, R. and Dermitzakis, E.T. (2010) Transcriptome genetics using second generation sequencing in a Caucasian population. Nature 464, 773–777.
[34] Zhang, K., Li, J.B., Gao, Y., Egli, D., Xie, B., Deng, J., Li, Z., Lee, J., Aach, J., Leproust, E.M., et al. (2009) Digital RNA allelotyping reveals tissue-specific and allele-specific gene expression in human. Nat. Methods 6, 613–618.
[35] Martinez-Canto, A., Castillejo, A., Mata-Balaguer, T., Castillejo, M.I., Hernandez-Illan, E., Irles, E., Barbera, V.M., Egoavil, C., Guarinos, C., Alenda, C., et al. (2012) TGFBR1 intralocus epistatic interaction as a risk factor for colorectal cancer. PLoS One 7, e30812.
[36] Taniguchi, Y., Choi, P.J., Li, G., Chen, H., Babu, M., Hearn, J., Emili, A. and Xie, X.S. (2010) Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. Science 329, 533–538.
[37] Crick, F.H.C. (1958) On protein synthesis. Symp. Soc. Exp. Biol. XII, 139–163.
[38] Lewin, R. (1983) Editing mRNA precursors. Science 222, 493.
[39] Li, M., Wang, I.X., Li, Y., Bruzel, A., Richards, A.L., Toung, J.M. and Cheung, V.G. (2011) Widespread RNA and DNA sequence differences in the human transcriptome. Science 333, 53–58.
[40] Pickrell, J.K., Gilad, Y. and Pritchard, J.K. (2012) Comment on "Widespread RNA and DNA sequence differences in the human transcriptome". Science 335, 1302.
[41] Hayden, E.C. (2011) Evidence of altered RNA stirs debate. Nature 473, 432.
[42] Schrider, D.R., Gout, J.F. and Hahn, M.W. (2011) Very few RNA and DNA sequence differences in the human transcriptome. PLoS One 6, e25842.
[43] Barraud, P. and Allain, F.H. (2012) ADAR proteins: double-stranded RNA and Z-DNA binding domains. Curr. Top. Microbiol. Immunol. 353, 35–60.
[44] Bahn, J.H., Lee, J., Li, G., Greer, C., Peng, G. and Xiao, X. (2012) Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. Genome Res. 22, 142–150.
[45] Peng, Z., Cheng, Y., Tan, B.C., Kang, L., Tian, Z., Zhu, Y., Zhang, W., Liang, Y., Hu, X. and Tan, X. (2012) Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. Nat. Biotechnol. 30, 253–260.
[46] Chen, R., Mias, G.I., Li-Pook-Than, J., Jiang, L., Lam, H.Y.K., Chen, R., Miriami, E., Karczewski, K.J., Hariharan, M., Dewey, F.E., et al. (2012) Personal omics profiling reveals dynamic molecular and medical phenotypes. Cell 148, 1293–1307.
[47] Taddei, F., Hayakawa, H., Bouton, M.F., Cirinesi, A.M., Matic, I., Sekiguchi, M. and Radman, M. (1997) Counteraction by MutT protein of transcriptional errors caused by oxidative damage. Science 278, 128–130.
[48] Kong, Q. and Lin, C.G. (2010) Oxidative damage to RNA: mechanisms, consequences, and diseases. Cell. Mol. Life Sci. 67, 1817–1829.
[49] Karijolich, J. and Yu, Y.T. (2011) Converting nonsense codons into sense codons by targeted pseudouridylation. Nature 474, 395–398.
[50] Rodnina, M.V. and Wintermeyer, W. (2001) Ribosome fidelity: tRNA discrimination, proofreading and induced fit. Trends Biochem. Sci. 26, 124–130.
[51] Cochella, L. and Green, R. (2005) Fidelity in protein synthesis. Curr. Biol. 15, R536–R540.
[52] Fersht, A.R. (1977) Editing mechanisms in protein synthesis. Rejection of valine by the isoleucyl-tRNA synthetase. Biochemistry 16, 1025–1030.
[53] Guo, M. and Schimmel, P. (2012) Structural analyses clarify the complex control of mistranslation by tRNA synthetases. Curr. Opin. Struct. Biol. 22, 119–126.
[54] Ogle, J.M. and Ramakrishnan, V. (2005) Structural insights into translational fidelity. Annu. Rev. Biochem. 74, 129–177.
[55] Pape, T., Wintermeyer, W. and Rodnina, M. (1999) Induced fit in initial selection and proofreading of aminoacyl-tRNA on the ribosome. EMBO J. 18, 3800–3807.
[56] Lee, J.W., Beebe, K., Nangle, L.A., Jang, J., Longo-Guess, C.M., Cook, S.A., Davisson, M.T., Sundberg, J.P., Schimmel, P. and Ackerman, S.L. (2006) Editing-defective tRNA synthetase causes protein misfolding and neurodegeneration. Nature 443, 50–55.
[57] Nangle, L.A., Motta, C.M. and Schimmel, P. (2006) Global effects of mistranslation from an editing defect in mammalian cells. Chem. Biol. 13, 1091–1100.
[58] Li, L., Boniecki, M.T., Jaffe, J.D., Imai, B.S., Yau, P.M., Luthey-Schulten, Z.A. and Martinis, S.A. (2011) Naturally occurring aminoacyl-tRNA synthetases editing-domain mutations that cause mistranslation in *Mycoplasma* parasites. Proc. Nat. Acad. Sci. U S A 108, 9378–9383.
[59] Kramer, E.B. and Farabaugh, P.J. (2007) The frequency of translational misreading errors in *E. coli* is largely determined by tRNA competition. RNA 13, 87–96.
[60] Drummond, D.A. and Wilke, C.O. (2008) Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. Cell 134, 341–352.